Abstract Title Page

Title:
Teacher Performance Pay Signals and Student Achievement: Are Signals Accurate, and How Well Do They Work?

Authors and Affiliations:

David Manzeske, American Institutes for Research
Marshall Garland, Gibson Consulting Group
Ryan Williams, American Institutes for Research
Benjamin West, American Institutes for Research
Alexandra Manzella Kistner, American Institutes for Research
Amie Rapaport, Gibson Consulting Group

Abstract Body

Background / Context: *Description of prior research and its intellectual context.*

High-performing teachers tend to seek out positions at more affluent or academically challenging schools, which tend to hire more experienced, effective educators. Consequently, low-income and minority students are more likely to attend schools with less experienced and less effective educators (see, for example, DeMonte & Hanna, 2014; Office for Civil Rights, 2014). This trend is problematic as a result of the established link between educator quality and student achievement.

Since 2001, federal policy has required states to have equity plans "to ensure that poor and minority children are not taught at higher rates than other children by inexperienced, unqualified, or out-of-field teachers" (U.S. Department of Education, 2002). On June 1, 2015, state education agencies submitted updated equitable access plans to the U.S. Department of Education (ED).

A potential lever that states and school districts have used to ensure the equitable distribution of effective teachers in high-need schools is to provide financial incentives. The goals of these incentives are to retain (or attract) the highest performing teachers and to motivate lower performers to enhance their practice to improve student achievement in high-need schools.

The premise of providing performance pay is that a financial incentive sends a signal to the labor force. Those who obtain the bonus receive a signal that their performance meets or exceeds expectations. In effect, they are told to "keep up the good work." Those who do not receive a bonus are sent a different message: their performance is unsatisfactory, and remediation is necessary. It is possible that when these signals are transmitted, high-performing teachers are motivated by the incentive—i.e., to maintain their performance level and to remain in their school—and lower performing teachers are motivated (or required) either to improve or to leave the profession.

Purpose / Objective / Research Question / Focus of Study: *Description of the focus of the research.*

In 2006, Congress established the Teacher Incentive Fund (TIF) to authorize ED to issue competitive grants to school districts and other education agencies. Goals of the TIF program include implementing performance-based compensation systems in high-need schools, increasing the number of effective teachers who are teaching disadvantaged students, and improving student achievement.

In partnership with a TIF grantee and 12 school districts, the purpose of this multiyear study is to examine local grant implementation and to investigate summative questions such as the following:

1. What percentage of teachers receives performance-based bonuses based on observations of classroom practice and value-added measures of student growth?
2. How do bonuses differ according to the performance levels of each performance measure (i.e., classroom practice, value-added score)?

3. How equitably are bonuses distributed to teachers within each performance measure, as well as between the two performance measures (i.e., classroom practice, value-added score)?
4. How do teacher retention rates differ according to the performance levels of each performance measure (i.e., classroom practice, value-added score)?
5. What is the impact of TIF program initiatives on student achievement?

Panelists will share considerations for policymakers who seek to implement similar initiatives and for researchers who seek to help others understand how teacher incentive policies function. See the Findings/Results section later in this document for the proposed panel plan (e.g., presentation order, audience participation).

Setting: *Description of the research location.*

The study takes place in schools serving students in Grades K–12 within a southern state. The schools had among the highest levels of student poverty and the lowest levels of student performance in their respective school districts.

Population / Participants / Subjects: *Description of the participants in the study.*

Participants include regular instructional, classroom teachers in 37 schools across 12 school districts. Each study year, approximately 1,000 teachers and their students were included in the study, as well as a population of comparison students from nonparticipating schools throughout the same school districts and throughout the state.

Intervention / Program / Practice: *Description of the intervention, program, or practice.*

From the 2011–12 through the 2013–14 school years, high-need schools received TIF funding to implement a compensation system based on two performance measures: observations of classroom practice and value-added measures of student growth (which were delivered to teachers during the following school year). Classroom practice scores ranged from 1.0 to 5.0 and were collected by trained evaluators during the school year. A minimum score of 2.5 was required for teachers to obtain a bonus. Teachers in tested grades and subjects received a value-added score that ranged between 1.0 and 5.0. A minimum score of 3.0 was required for teachers to obtain a bonus.[1] All teachers were eligible for schoolwide awards based on schoolwide, value-added scores. In addition to the above performance feedback, classroom teachers received support from mentor and master teachers, and teachers participated in teaming activities to improve their practice.

Research Design: *Description of the research design.*

To answer research questions (RQs) 1 through 4, data were used from the 2011–12 through the 2013–14 school years. Within-year comparisons were conducted for TIF school teachers. To answer RQ5, a comparative interrupted time series (CITS) design was implemented using data

[1] Bonus amounts were not discrete: teachers with performance scores greater than the minimum score received larger bonuses.

from the 2002–03 through the 2013–14 school years. A comparison group description is located in the Findings / Results section.

Data Collection and Analysis: *Description of the methods for collecting and analyzing data.*

Data collection. Archival records were used to answer the research questions listed earlier. Teacher-level records include grant-generated classroom practice and value-added performance ratings, grant-specific bonus amounts, and district administrative records for teachers in TIF schools. Statewide student achievement data were publicly available and aggregated at each grade level within each school and within each school year.

Analysis. A variety of descriptive and inferential analyses were performed. To answer RQ1 and RQ2, descriptive analyses were performed. To answer RQ3, the Gini coefficient was calculated for each bonus source to measure and compare the degree of disbursement equality for each performance standard. The Gini coefficient is used to measure income inequality, and it has been used in other evaluations of performance pay plans to assess how evenly (or unevenly) performance bonuses are distributed within and across participating schools (see Springer et al., 2010). The coefficient range is between 0 and 1: 0 reflects perfect equality (meaning that the bonuses were evenly disbursed across a large pool of eligible participants), whereas 1 represents perfect inequality (meaning that a small number of eligible participants received relatively large payouts). For instance, if a school awarded its entire bonus allocation to a single teacher, then the Gini coefficient would be 1. If every teacher at a school received precisely the same incentive, then the Gini coefficient would be 0. To answer RQ4, predicted probabilities were derived from a logistic regression model, with centered teacher- and school-level covariates (e.g., teacher's years of experience, percentage of English learners). To answer RQ5, a CITS analysis was performed using nine years of preinterruption data. In the main, or confirmatory, model, the research team first compared outcomes at TIF schools with those in all nonparticipating schools in TIF districts (N = 138). In addition, the research team conducted three sensitivity tests in which selection criteria for the comparison group were altered (e.g., using propensity scores).

Findings / Results: *Description of the main findings with specific details.*

Two panelists will present the study findings after an introduction to the relevant policy context and the study background from the moderator, David Manzeske of American Institutes for Research (AIR). The first panelist, Marshall Garland of Gibson Consulting Group, will present findings related to the teacher performance measures, bonuses, and retention. The second panelist, Ryan Williams of AIR, will present student achievement impact findings. Time will be allocated for audience participation. The proposed plan for the panel is as follows:

- Policy context and study background (5 minutes)
- Findings related to teacher performance measures, bonuses, and retention (25 minutes)
- Audience comments and questions (20 minutes)
- Findings related to impacts on student achievement (20 minutes)
- Audience comments and questions (20 minutes)

Findings correspond to program implementation during the 2011–12, 2012–13, and 2013–14 school years.[2] These findings are summarized in the list that follows.

- **Each year, nearly every teacher received a performance bonus**. Between 97 percent and 99 percent of career teachers received an annual bonus based on their classroom practice score (i.e., a score of at least 2.5 of 5.0). Between 59 and 64 percent of career teachers received an annual bonus based on their classroom value-added score (i.e., a score of at least 3.0 of 5.0). Between 36 percent and 60 percent of career teachers received an annual bonus based on their school value-added score. See Figure 1.
- **Average annual bonuses were larger among those who received a higher performance rating.** Average classroom practice bonuses ranged between $485 and $2,032 (for scores of 2.5 to 5.0, respectively). Average classroom value-added bonuses ranged between $443 and $2,671 (for scores of 3.0 to 5.0, respectively). See Figures 2 and 3.
- **Bonuses based on classroom practice scores were more evenly disbursed across the eligible teachers and performance ratings than bonuses based on value-added scores.** Among career teachers, the mean Gini coefficient was 0.19 for classroom practice performance and 0.51 for value-added performance. See Figure 4.
- **The probability of returning to the same school is positively associated with classroom practice ratings, whereas the association between value-added ratings and retention is mostly flat.** For classroom practice ratings, the largest difference in the probability of school retention is between ratings that are just below the bonus cutoff and those that are just above it; on average, the ratings jump approximately 25 percentage points. Above the bonus cutoff, the probability of retention weakens as classroom practice scores increase. See Figure 5.
- **The study found no statistically significant effects on student mathematics or reading achievement that were robust to multiple specifications.** A series of analyses were performed using one main analysis for each outcome and three sensitivity analyses for each outcome. The sensitivity analyses were designed to assess the extent to which results varied based on selection criteria for the comparison group. See Table 1.

Conclusions: *Description of conclusions, recommendations, and limitations based on findings.*

The TIF grantee implemented a performance-based compensation system in 37 schools across 12 districts. Teachers with higher performance ratings received larger bonus amounts, and all teachers received professional support to improve their practice; however, the study found no impacts on student achievement. Any of several explanations could account for these null findings. First, because nearly every teacher received a performance-based bonus, it is possible that the performance bonus signal was misapplied to the lower performing teachers. Arguably, it is most important for the lower performers to improve their practice, yet it is possible that their bonus (although relatively smaller) was interpreted as a message that their performance was sufficient. Second, although higher performing teachers received larger bonuses, it is possible that their lower performing counterparts did not believe that the marginally larger, average payout amount was worth the effort that would be required to obtain a higher performance score

[2] Analysis of data from the 2014–15 school year is ongoing.

(and, thus, a larger payout). Third, with the exception of the few teachers who received a classroom practice score that was too low for them to obtain a bonus, the probability of retention did not differ markedly for teachers with varied performance scores. Therefore, whereas it is important that the higher performing teachers be retained, it is possible that the lower performing teachers did not choose to leave the profession because they believed that their performance was sufficient; thus, there were fewer opportunities to replace them with more effective teachers. Taken together, it is possible that the performance pay signals were applied too liberally (to teachers who are not yet performing at high enough levels to improve student achievement), that the signals were not strong enough (in bonus amounts) to produce marked differences in teacher practice, or that both liberally applied performance pay signals and insufficiently strong signals were factors in the lack of findings on the impacts upon student achievement.

When practitioners and policymakers choose to implement performance-based compensation systems, the direction and strength of the performance incentives ought to be considered. Moreover, checks of the signals—to whom they are applied and the amounts that are offered—are necessary. More research is required, however, to inform policy and program development with respect to context-specific bonus amounts that will motivate changes in teacher practice.

The generalizability of study findings is limited to schools and school districts that tend to be smaller and more rural.

Appendices

Appendix A. References

References are to be in APA version 6 format.

DeMonte, J., & Hanna, R. (2014). *Looking at the best teachers and who they teach: Poor students and students of color are less likely to get highly effective teaching*. Washington, DC: Center for American Progress. Retrieved from https://www.americanprogress.org/wp-content/uploads/2014/04/TeacherDistributionBrief1.pdf

Office for Civil Rights, U.S. Department of Education. (2014). *Civil rights data collection: Data snapshot; Teacher equity* (Issue Brief No. 4). Washington, DC: Author. Retrieved from http://www2.ed.gov/about/offices/list/ocr/docs/crdc-teacher-equity-snapshot.pdf

Springer, M. G., Lewis, J. L., Ehlert, M. W., Podgursky, M. J., Crader, G. D., Taylor, L. L., et al. (2010). District Awards for Teacher Excellence (D.A.T.E.) program: Final evaluation report [NCPI Policy Evaluation Report]. Nashville, TN: Vanderbilt University, National Center on Performance Incentives.

U.S. Department of Education. (2015). *Laws and guidance: Elementary and secondary education, part A, subpart 1, sec. 1111 et seq.* Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/pg2.html

Appendix B. Tables and Figures

Figure 1. Percentage of Teachers Who Received Performance Awards by Performance Measure, 2011–12 to 2013–14

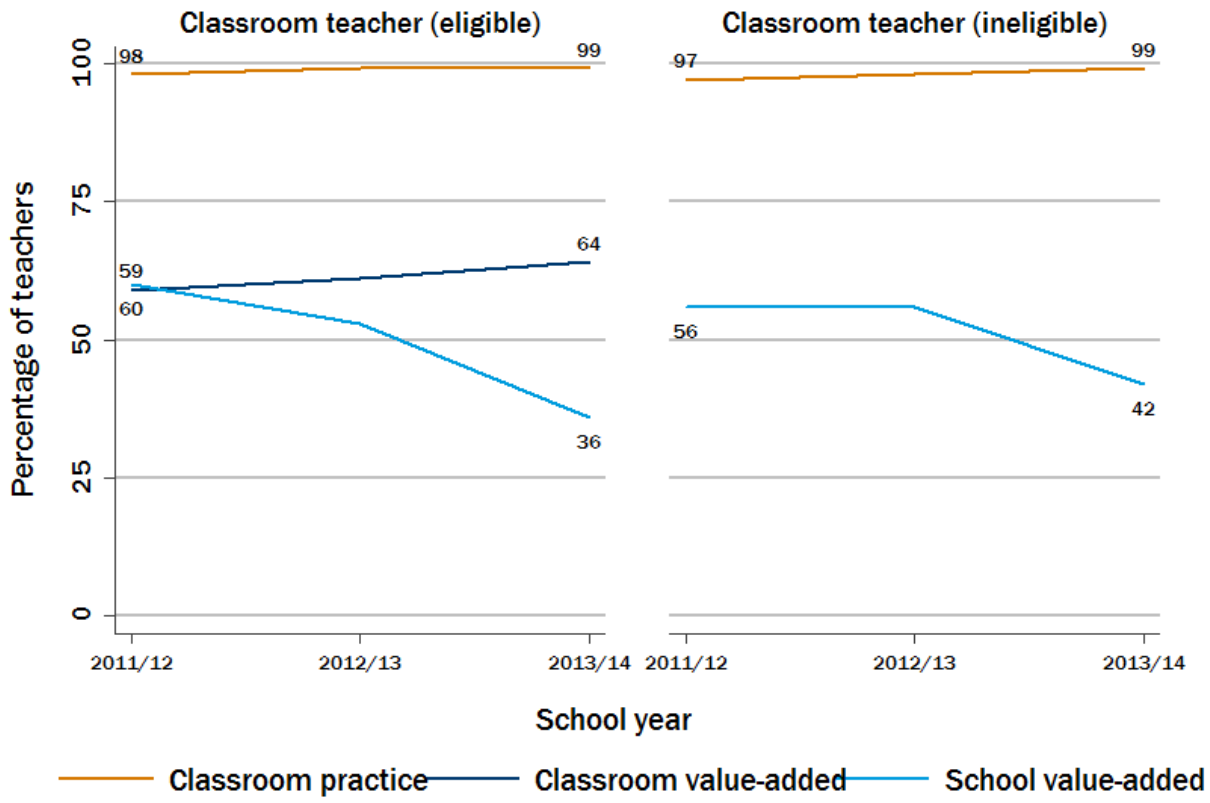
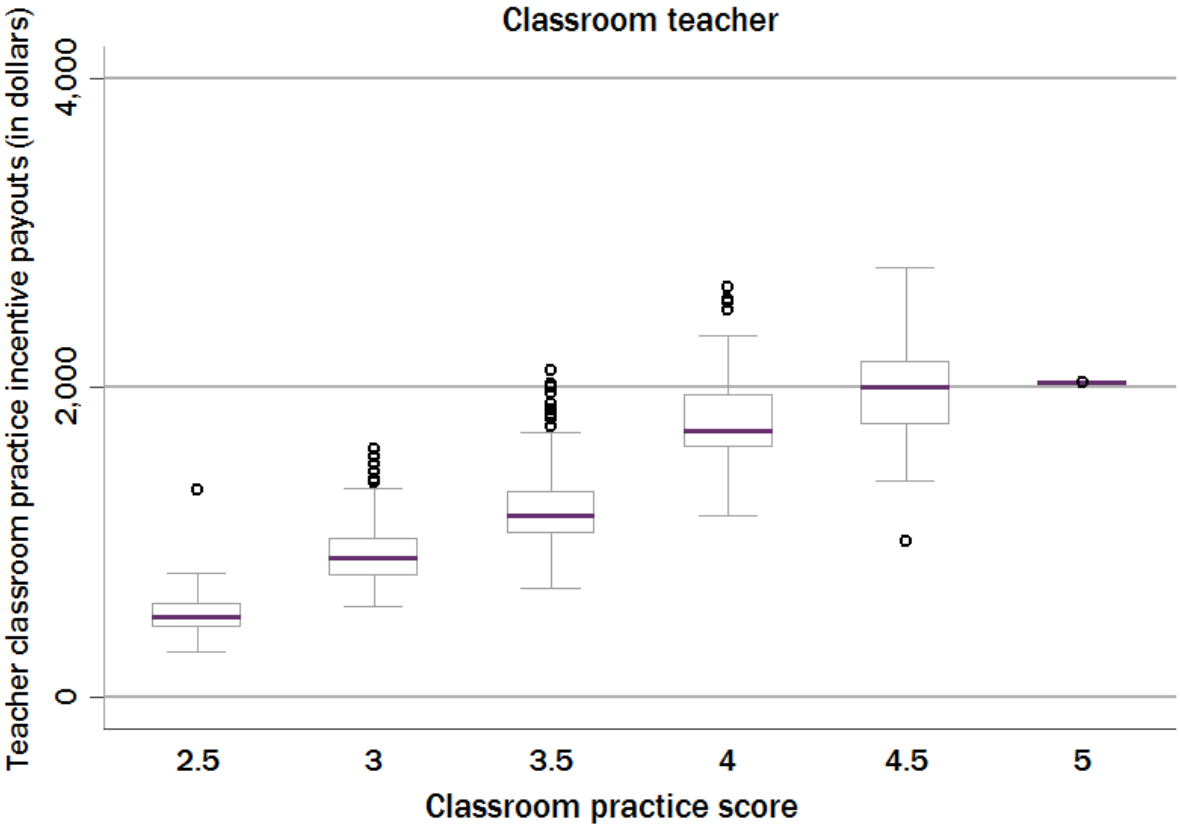
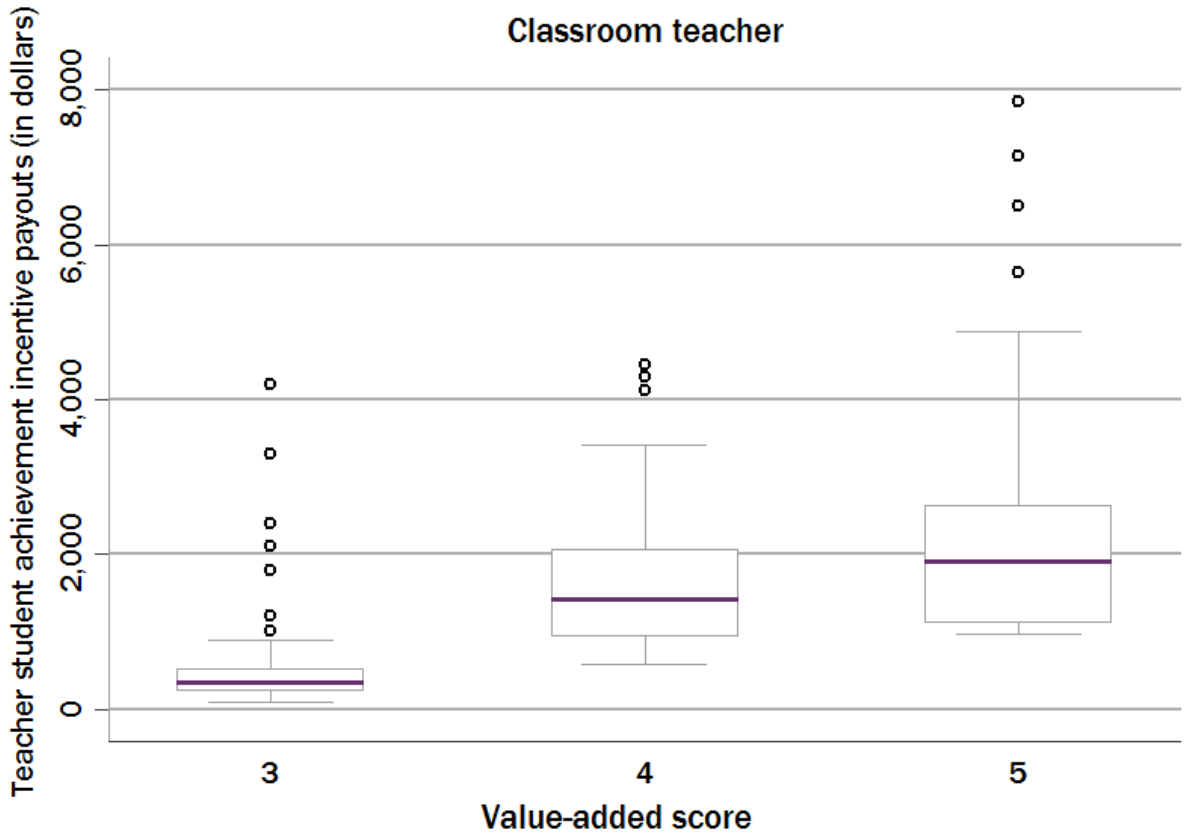


Figure 2. Distribution of Classroom Practice Bonuses and Classroom Practice Score, 2011–12 to 2013–14



Note. Classroom teachers must achieve a minimum score of 2.5 or higher to obtain a bonus.

Figure 3. Distribution of Classroom Achievement Incentives and Value-Added Score, 2011–12 to 2013–14



Note. Classroom teachers must achieve a minimum score of 3.0 or higher to obtain a bonus.

Figure 4. Mean Gini Coefficient, by Incentive Source, 2011–12 to 2013–14

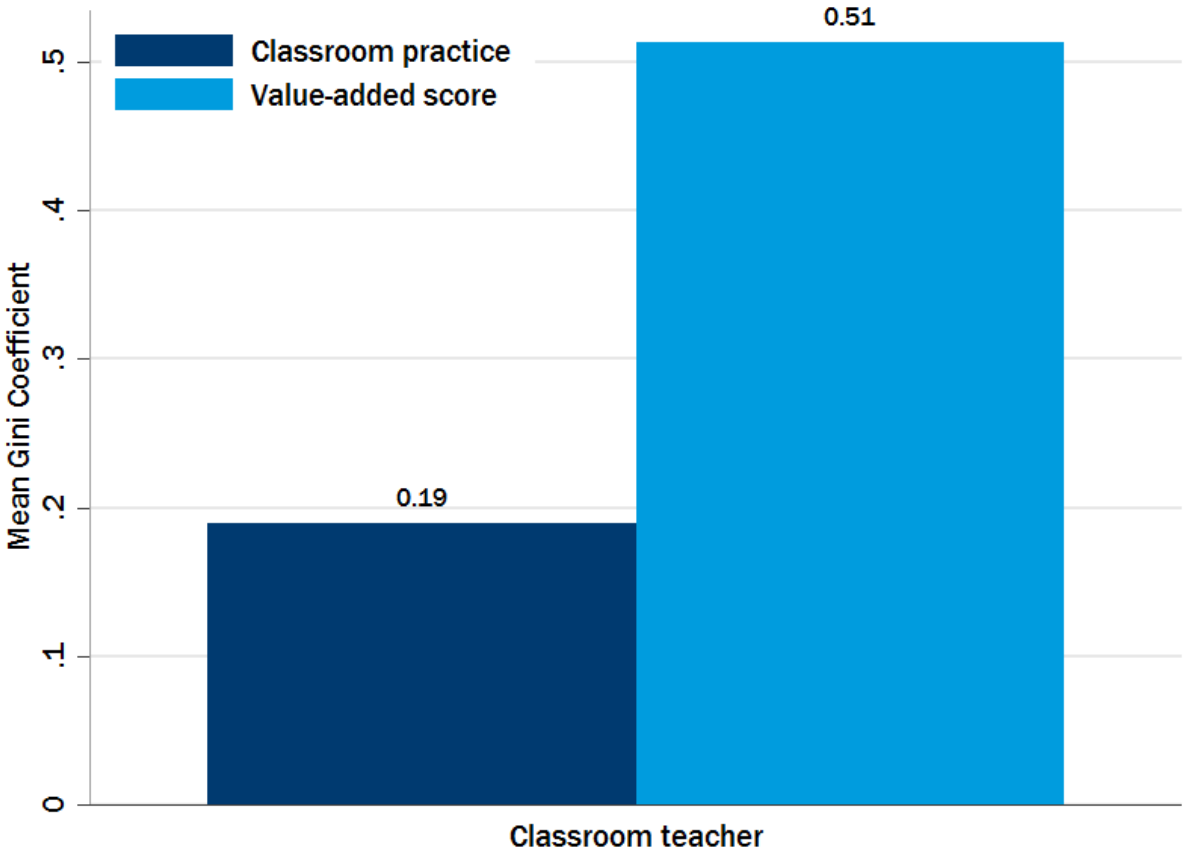
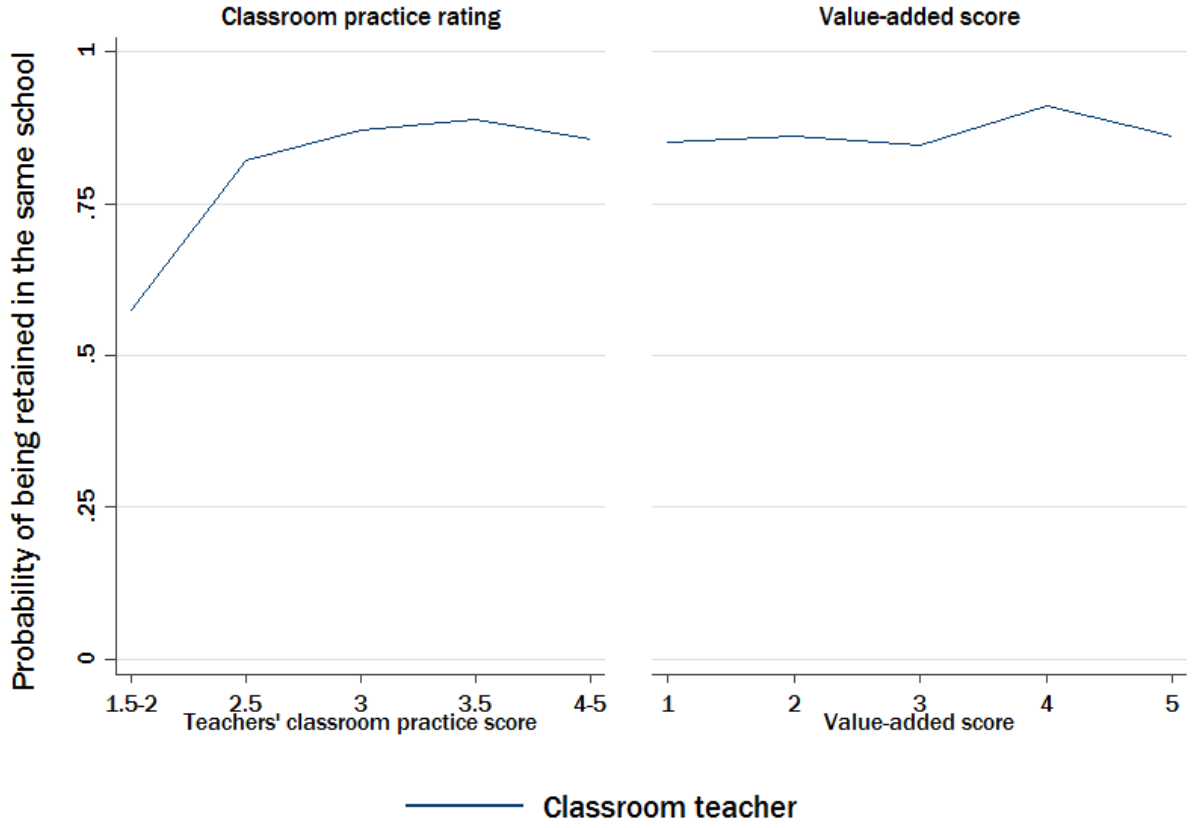


Figure 5. Predicted Probabilities of Remaining at the Same School Derived From a Logistic Regression, by Teacher Classroom Practice and Value-Added Scores, 2011–12 to 2014–15



Notes: Predicted probabilities were derived from a logistic regression, with covariates held at their mean values. Covariates included in the model encompass the following: Teacher covariates are total number of years of professional experience, highest degree level attained, ethnicity, school year, and sex; and school covariates are percentage of English learners, number of students enrolled, and school level (e.g., elementary school, middle school, or high school); percentage of economically disadvantaged students; and percentage of minority students (summed percentage of Hispanic and Black students).

Table 1. Impact Estimates for Groups A, B, C, and D

Subject		Reading				Mathematics			
Measure		**Group A**	**Group B**	**Group C**	**Group D**	**Group A**	**Group B**	**Group C**	**Group D**
Year 1	Estimate	0.03	-0.08	-0.12	-0.12	0.06	-0.14	-0.16	-0.21
	SE	(0.08)	(0.09)	(0.09)	(0.10)	(0.11)	(0.12)	(0.12)	(0.13)
Year 2	Estimate	0.10	0.01	-0.05	-0.06	0.26*	0.07	-0.02	-0.09
	SE	(0.08)	(0.10)	(0.09)	(0.10)	(0.10)	(0.12)	(0.12)	(0.13)
Year 3	Estimate	0.18*	0.08	0.01	0.00	0.20	0.01	-0.06	-0.12
	SE	(0.09)	(0.10)	(0.10)	(0.11)	(0.11)	(0.12)	(0.13)	(0.14)

Notes. SE = standard error. Standard errors appear in parentheses. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.